



Data Article

Dataset of foraminiferal sedimentary DNA (*sedDNA*) sequences from Svalbard



Joanna Pawłowska^{a,*}, Jan Pawłowski^{a,b}, Marek Zajączkowski^a

^a Institute of Oceanology Polish Academy of Sciences, Sopot 81-712, Poland

^b Department of Genetics and Evolution, University of Geneva, Geneva CH 1211, Switzerland

ARTICLE INFO

Article history:

Received 5 March 2020

Accepted 2 April 2020

Available online 18 April 2020

Keywords:

Metabarcoding
sedimentary DNA
Environmental DNA
Foraminifera
Svalbard

ABSTRACT

Environmental DNA (eDNA) is usually defined as genetic material obtained directly from environmental samples, such as soil, water, or ice. Coupled to DNA metabarcoding, eDNA is a powerful tool in biodiversity assessments. Results from eDNA approach provided valuable insights to the studies of past and contemporary biodiversity in terrestrial and aquatic environments. However, the state and fate of eDNA are still investigated and the knowledge about the form of eDNA (i.e., extracellular vs. intracellular) or the DNA degradation under different environmental conditions is limited. Here, we tackle this issue by analyzing foraminiferal sedimentary DNA (*sedDNA*) from different size fractions of marine sediments: >500 μm , 500–100 μm , 100–63 μm , and < 63 μm . Surface sediment samples were collected at 15 sampling stations located in the Svalbard archipelago. Sequences of the foraminifera-specific 37f region were generated using Illumina technology. The presented data may be used as a reference for a wide range of eDNA-based studies, including biomonitoring and biodiversity assessments across time and space.

© 2020 The Authors. Published by Elsevier Inc.

This is an open access article under the CC BY license.
(<http://creativecommons.org/licenses/by/4.0/>)

* Corresponding author.

E-mail address: pawlowska@iopan.pl (J. Pawłowska).

Specification table

Subject	Environmental Science, Oceanography, Molecular Biology,
Specific subject area	Foraminifera biodiversity and spatial distribution
Type of data	Tables
How data were acquired	Amplicon high-throughput sequencing with Illumina technology. SLIM pipeline to produce OTU-to-sample tables.
Data format	Analyzed data
Parameters for data collection	Surface sediment samples collected with box corer during the cruise of R/V Oceania in 2016.
Description of data collection	The surface sediment samples were collected at 15 sampling stations located in the Svalbard Archipelago. The total DNA was extracted from sediments, foraminifer-specific DNA fragments were amplified and sequenced according to standard protocols.
Data source location	Institute of Oceanology Polish Academy of Sciences, Sopot, Poland
Data accessibility	Repository name: Mendeley Data Data identification number: doi: 10.17632/7kjkf8by5d.1 Direct URL to data: https://data.mendeley.com/datasets/7kjkf8by5d/1

Value of the data

- The data provides the first insight into the genetic diversity of Arctic foraminifera in different sediment size fractions.
- Also, it's an overview of the spatial distribution of Arctic foraminifera inferred from *sedDNA*.
- This data may serve as a reference in a wide range of metagenomics-based studies, including biomonitoring, biodiversity surveys, and environmental impact assessment studies.
- The data may be used also in the studies of past climatic and environmental changes.

1. Data description

The dataset contains foraminiferal *sedDNA* sequences from four sediments size fractions: >500 μm , 500–100 μm , 100–63 μm , and <63 μm . Sequences are clustered into the Operational Taxonomic Units (OTUs), and for each OTU, the number of sequence reads is presented. The data set can be accessed at Mendeley Data (doi: 10.17632/7kjkf8by5d.1). The samples were collected at 16 sampling stations collected from five localities in the Svalbard archipelago (Fig. 1). Sampling stations coordinates and sampling depths can be found in Table 1. The total number of OTUs recorded in each sampling location is presented in Fig. 2. The number and percentage

Table 1

Geographical coordinates and depths of sampling stations.

Station	Depth [m]	Latitude [°N]	Longitude [°W]
Isfjorden 1	191	78 24.000	15 35.690
Isfjorden 2	240	78 15.890	14 51.200
Wijdefjorden 1	160	79 53.815	15 19.066
Wijdefjorden 3	154	79 09.464	16 00.085
Rijpfjorden 1	202	80 05.374	22 12.966
Rijpfjorden 2	260	80 18.464	22 10.951
Rijpfjorden 3	260	80 22.354	22 05.736
Rijpfjorden 4	155	80 30.288	22 02.996
Nordauslandet 4	42	79 42.727	26 34.955
Nordauslandet 5	82	79 40.506	26 48.952
Nordauslandet 6	147	79 36.241	27 30.196
Nordauslandet 7	275	79 33.185	28 00.347
Edgeøya 1	41	77 42.925	24 12.853
Edgeøya 2	71	77 41.147	24 39.384
Edgeøya 3	114	77 39.751	25 00.162

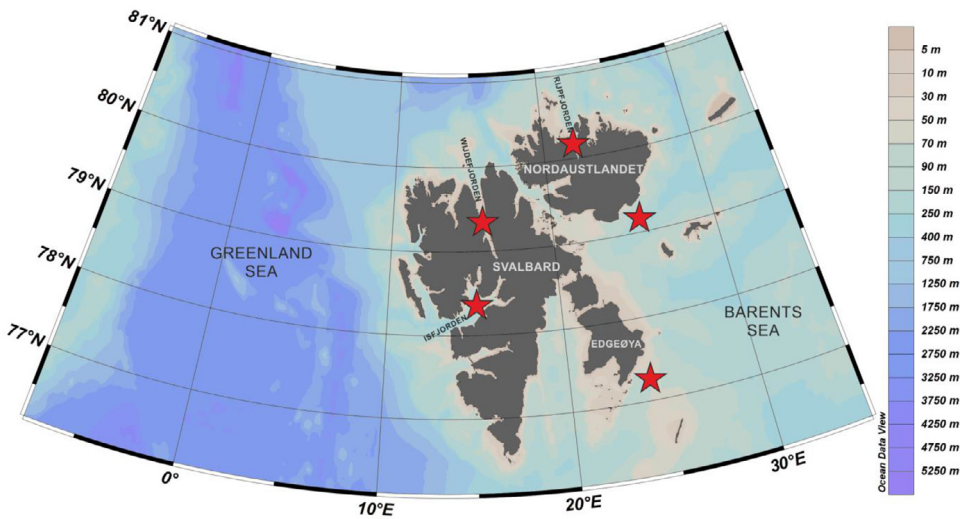


Fig. 1. Study area and sampling localities.

of OTUs and the percentage of DNA sequences found in the certain sediment size fractions are presented on the Venn diagrams (Fig. 3).

2. Experimental design, materials, and methods

2.1. Sampling

Surface sediment samples were collected with the use of box corer during the cruise of R/V Oceania in August 2016. The upper 2 cm of sediment has been sampled from the surface of approximately 25 cm². Samples for sedimentary DNA (*sedDNA*) analysis were wet sieved on a 500 μm, 100 μm, and 63 μm sieves. A fraction smaller than 63 μm was retained. Samples were transferred to sterile containers and froze in −20 °C.

2.2. *sedDNA* analysis

The DNA from sediment fractions 500 μm, 100 μm, and 63 μm was extracted from 0.25 g of bulk sediment with DNeasy PowerSoil Kit (Qiagen). Due to a large amount of sediment in <63 μm fraction, DNA was extracted from 10 g of sediment using DNeasy PowerMax Soil Kit (Qiagen). The SSU DNA fragment including foraminifera-specific 37f hypervariable region [1] has been PCR amplified with the s14F1 (5'-XXXXXAAGGGCACCACAAGAACGC-3') and s15 (5'-XXXXXCCTATCACATAATCATGAAAG-3') primers tagged with unique sequences of 5 nucleotides appended at their 5' ends. For each sample, 3 PCR replicates were prepared. Amplicons were quantified with Qubit 3.0 fluorometer and the pool was purified with High PCR Cleanup Micro Kit (Roche). Library preparation was performed with TruSeq DNA PCR-Free LT Library Prep Kit (Illumina) and was loaded onto a MiSeq instrument for a paired-end HTS run of 2 × 150 cycles.

2.3. Post-sequencing data processing

Raw sequence data were processed according to [2,3]. The post-sequencing data processing was performed using the SLIM web app [4] and included demultiplexing the libraries, join-

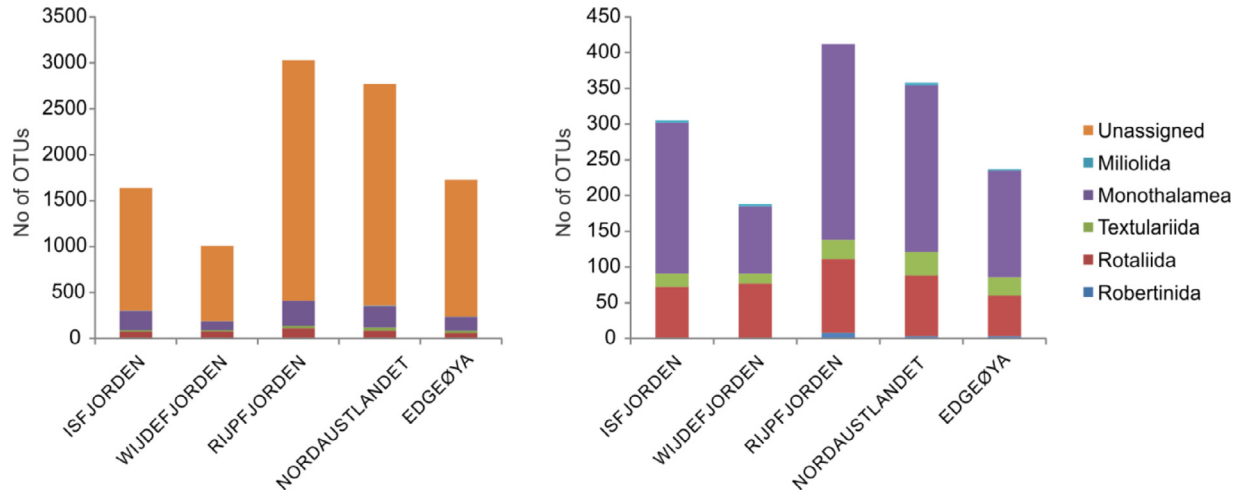


Fig. 2. Number of OTUs found in each sampling locality, including unassigned OTUs (left panel) and excluding unassigned OTUs (right panel).

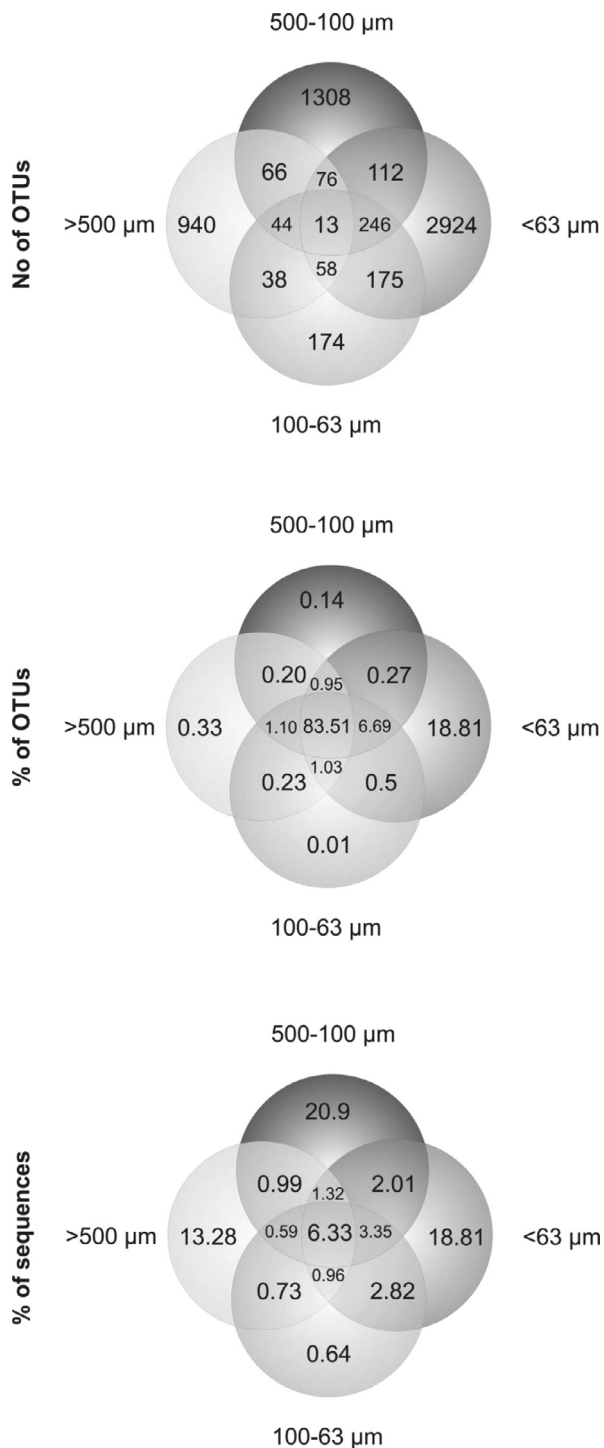


Fig. 3. Venn diagrams of number of OTUs, percent [%] of OTUs, and percent [%] of sequences found in the studied sediment fractions (>500 μm, 500–100 μm, 100–63 μm, <63 μm).

ing the paired-end reads, chimera removal, Operational Taxonomic Units (OTUs) clustering, and taxonomic assignment. Sequences were clustered into OTUs using the Swarm module [5] and each OTU was assigned to the highest possible taxonomic level using vsearch [6] against a local database of foraminiferal SSU DNA sequences. The results were presented as OTUs-to-samples tables.

Acknowledgments

Mz and Joanna P were supported by the National Science center in Poland projects no. 2016/21/B/ST10/02308 and 2018/31/B/ST10/01616, respectively. Jan P was supported by Swiss National Science Foundation project no. 31003A_179125.

Authors contributions

Joanna P, Jan P and MZ formulated the research goals and developed the research methodology. Joanna P performed laboratory analysis. Joanna P, Jan P and MZ analyzed and interpret the data. Joanna P wrote the manuscript.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.dib.2020.105553](https://doi.org/10.1016/j.dib.2020.105553).

References

- [1] J. Pawłowski, B. Lecroq, Short rDNA barcodes for species identification in foraminifera, *J. Eukaryot. Microbiol.* 57 (2010) 197–205, doi:[10.1111/j.1550-7408.2009.00468.x](https://doi.org/10.1111/j.1550-7408.2009.00468.x).
- [2] J. Pawłowska, F. Lejzerowicz, P. Esling, W. Szczuciński, M. Zajączkowski, J. Pawłowski, Ancient DNA sheds new light on the Svalbard foraminiferal fossil record of the last millennium, *Geobiology* 12 (2014) 277–288, doi:[10.1111/gbi.12087](https://doi.org/10.1111/gbi.12087).
- [3] J. Pawłowska, M. Zajączkowski, M. Łacka, F. Lejzerowicz, P. Esling, J. Pawłowski, Palaeoceanographic changes in horn-sund fjord (Spitsbergen, Svalbard) over the last millennium: new insights from ancient DNA, *Clim. Past* 12 (2016) 1459–1472, doi:[10.5194/cp-12-1459-2016](https://doi.org/10.5194/cp-12-1459-2016).
- [4] Y. Dufresne, F. Lejzerowicz, L. Apotheloz Perret-Gentil, J. Pawłowski, T. Cordier, SLIM a flexible web application for the reproducible processing of environmental dna metabarcoding data, *BMC Bioinf.* 20 (2019) 88, doi:[10.1186/s12859-019-2663-2](https://doi.org/10.1186/s12859-019-2663-2).
- [5] F. Mahé, T. Rognes, C. Quince, C. de Vargas, M. Dunthorn, Swarm: robust and fast clustering method for amplicon-based studies, *Peer J.* 2 (2014) e593, doi:[10.7717/peerj.593](https://doi.org/10.7717/peerj.593).
- [6] T. Rognes, T. Flouri, B. Nichols, C. Quince, F. Mahé, VSEARCH: a versatile open source tool for metagenomics, *Peer J.* 4 (2016) e2584, doi:[10.7717/peerj.2584](https://doi.org/10.7717/peerj.2584).